Quantile Regression Learning with Coefficient Dependent l^q-Regularizer

Pengju Sun[1,*], *Meng* Li[2], and *Hongwei* Sun[1]

[1]School of Mathematical Sciences, University of Jinan, People's Republic of China
[2]Caoxian No.1 Senior High School ,People's Republic of China

Abstract. In this paper, We focus on conditional quantile regression learning algorithms based on the pinball loss and l^q-regularizer with 1≤q≤2. Our main goal is to study the consistency of this kind of regularized quantile regression learning. With concentration inequality and operator decomposition techniques, we obtained satisfied error bounds and convergence rates .

1INTRODUCTION

Let X be a compact subset of $\Re^n$ and Y=$\Re$, ρ be a Borel probability distribution on Z=X×Y which governs the relationship between the input data x and the response y. The set of samples $z=\{(x_i, y_i)\}_{i=1}^m$ are drawn independently from ρ . The goal of quantile regression learning is to estimate the conditional τ -quantile function $f_{\rho,\tau}$, it's value $f_{\rho,\tau}(x)$ is defined to be the τ -quantile of conditional distribution $\rho(\cdot|x)$, i.e., $f_{\rho,\tau}(x)=u$ such that

$$\rho(\{y\in(-\infty,u]\}|x)\ge\tau \text{ and } \rho(\{y\in(u,\infty]\}|x)\ge 1-\tau, x\in X \qquad (1)$$

Here $\tau\in(0,1)$ is a fixed constant specifying the desired quantile level. Throughout this paper, we assume that the conditional distribution $\rho(\cdot|x)$ is supported on [-1,1] almost surely for almost all $x\in X$. This assumption yields that $|f_{\rho,\tau}(x)|\le 1$ almost everywhere.

Since least square loss $\varphi(y,t)=(y-t)^2$ is appropriate for regression learning, and hinge loss $\varphi(y,t)=\max\{1-yt,0\}$ is for support vector machine for classification, the τ-pinball loss $\psi_\tau:\Re\to\Re_+$,

$$\psi_\tau(u)=\begin{cases}(1-\tau)u, & \text{if } u>0,\\ -\tau u, & \text{if } u\le 0.\end{cases}$$

is used in quantile regression learning. The associated generalization risk for $f: X\to\Re$ is

$$\varepsilon_\tau(f):=\int_{X\times Y}\psi_\tau(f(x)-y))d\rho(z)$$

and $f_{\rho,\tau}$ minimizes the generalization risk $\varepsilon_{\tau,\rho}(f)$, see [12].

Let K be a Mercer kernel, i.e., a continuous, symmetric, and semi-definite function defined on X×X. The reproducing kernel Hilbert space H_K is defined to be the completion of span $\{K_x=K(\cdot,x): x\in X\}$ with the inner product $<K_x,K_y>=K(x,y)$ (see [1] for detail). Kernel based quantile regression learning schemes can be described as

$$f_{z,\lambda}=\arg\min_{f\in H_K}\{\frac{1}{m}\sum_{i=1}^m\psi_\tau(f(x_i)-y_i)+\lambda\|f\|_K^2\} \quad (2)$$

Here H_K is taken as the hypothesis space, and the penalty term is based on the norm of functions in H_K. The error bound and asymptotic convergence of this learning scheme have been discussed, see [18,12] and references therein.

To balance the approximation ability and sparsity of the algorithm (2), the ε-insensitive pinball loss $\psi_\tau^{(\varepsilon)}:\Re\to\Re_+$

$$\psi_\tau^{(\varepsilon)}(u)=\begin{cases}(1-\tau)(u-\varepsilon), & \text{if } u>\varepsilon,\\ -\tau(u+\varepsilon), & \text{if } u\le\varepsilon,\\ 0, & \text{otherwise.}\end{cases}$$

is used to replace the τ-pinball loss ψ_τ , and performance of the associated regularization scheme

$$f_{z,\lambda}^{(\varepsilon)}=\arg\min_{f\in H_{K,z}}\{\frac{1}{m}\sum_{i=1}^m\psi_\tau^{(\varepsilon)}(f(x_i)-y_i)+\lambda\|f\|_K^2\} \quad (3)$$

is studied in [19].

Now, we restrict our attention to coefficient-based regularization schemes in a data dependent hypothesis

* Corresponding author: ss_sunpj@ujn.edu.cn

space. The hypothesis space here is determined by a kernel function $K: X \times X \to R$ and the sample set **z**,

$$H_{K,z} = \left\{ \sum_{i=1}^{m} \alpha_i K(x_i, \cdot) : \alpha = (\alpha_1, \cdots, \alpha_m) \in \Re^m \right\}.$$

Here the kernel K is only asked to be uniformly bounded and continuous. Since we don't require a Hilbert or Banach space norm for functions in the hypothesis space, the penalty term $\lambda \| f \|_K^2$ is replaced by some norm of coefficient vector $\alpha \in \Re^m$ of function f.

This kind coefficient-based regularized network was first introduced by Vapnik [15] to design linear programming support vector machines. Taking this regularization scheme, one can more freely choose kernel function K and different norms of coefficient vector as the regularizer to fit the data in certain trend. Coefficient-based regression learning with l^q- norm for 1≤q≤2 can be defined as $f_z = f_{\alpha_z}$ where

$$\alpha_z = \arg\min_{\alpha \in R^m} \left\{ \frac{1}{m} \sum_{i=1}^{m} (y_i - f_\alpha(x_i))^2 + \lambda m^{q-1} \sum_{i=1}^{m} |\alpha|^q \right\} \quad (4)$$

Mathematical analysis of coefficient-based regression learning (4) has been established, include, framework of analysis for coefficient-based regression learning is proposed in [17] , for coefficient-based regression learning with l^1-norm penalty see [20,8,13], for coefficient-based regression learning with l^2-norm penalty see [7,14], also for l^q-norm penalty see [6].

In the quantile regression learning, Li and Sun takes more general kernels and ε-insensitive pinball loss with l^2-norm and l^1-norm regularizer respectively, see [4,5] .

In this paper, we consider the following quantile regression learning scheme with ε-insensitive pinball loss and coefficient-based l^q-norm regularization (1≤q≤2),

$$f_{z,\lambda}^{(\varepsilon)} = \arg\min_{f \in H_{K,z}} \{ \frac{1}{m} \sum_{i=1}^{m} \psi_\tau^{(\varepsilon)}(f(x_i) - y_i) + \lambda \Omega(f) \} \quad (5)$$

where $\Omega(f) = m^{q-1} \sum_{i=1}^{m} |\alpha|^q$ and $f = \sum_{i=1}^{m} \alpha_i K(\cdot, x_i)$. In the sequel, the empirical τ-quantile error of a function f is denoted by

$$\varepsilon_{\tau,z}(f) := \frac{1}{m} \sum_{i=1}^{m} \psi_\tau^{(\varepsilon)}(f(x_i) - y_i) \quad (6)$$

We will prove the asymptotic convergence of this learning scheme, i.e., how the output function $f_{z,\lambda}^{(\varepsilon)}$ approximates the quantile regression function $f_{\rho,\tau}$ as m→ ∞ . In fact, we extend the existing results of this learning schemes from l^1 and l^2 regularization to more general l^q regularization (1≤q≤2).

The rest of this paper is organized as follows. In section 2, we give the assumptions and main result. In Section 3, we give the estimates for hypothesis errors and sample errors. In Section 4 we give the error bound and learning rate by iteration method.

2 Assumptions and Main Result

The projection operator (see [2]) deals with heavy tailed noise well, that is helpful in obtaining our main result of $f_{z,\lambda}^{(\varepsilon)}$ approximating $f_{\rho,\tau}$.

Definition 1. The projection operator π on the space of function on X is defined by

$$\pi(f)(x) = \begin{cases} 1, & \text{if } f(x) > 1, \\ -1, & \text{if } f(x) < -1, \\ f(x), & \text{if } -1 \le f(x) \le 1. \end{cases}$$

Since $| f_{\rho,\tau}(x) | \le 1$ almost everywhere, $\pi(f_{z,\lambda}^{(\varepsilon)})$ is a better approximation of $f_{\rho,\tau}$ than $f_{z,\lambda}^{(\varepsilon)}$. We will bound the error of $\pi(f_{z,\lambda}^{(\varepsilon)})$ approximating $f_{\rho,\tau}$ in function space $L_{\rho_X}^{p^*}$. Here the index $p^* = \frac{pp'}{p+1}$, and p, p' are the parameters in the following distribution condition proposed in [10]. This p-average type p' condition can measure the learnability of τ-quantile, also guarantee the uniqueness of τ-quantile function $f_{\rho,\tau}$.

Definition 2. Let $p \in (0, +\infty]$ and $p' \in (1, +\infty)$. We say that ρ has a τ-quantile of p-average type p' if for almost all $x \in X$, there exist a τ-quantile $t^* \in \Re$ and constants $a_x \in (0,2], b_x > 0$ such that for each $s \in [0, a_x]$,

$$\rho(\{y \in (t^* - s, t^*)\} | x) \ge b_x s^{p'-1}, \text{ and}$$

$$\rho(\{y \in (t^*, t^* + s)\} | x) \ge b_x s^{p'-1}, \quad (7)$$

and that the function $\gamma: X \to [0, \infty]$, $\gamma(x) = b_x a_x^{p'-1}$, satisfies $\gamma^{-1} \in L_{\rho_X}^p$.

Under the p-average type p' condition with $p \in (0, +\infty]$ and $p' \in (1, +\infty)$. It is proved in [11] that for any measurable function f on X,

$$\| f - f_{\rho,\tau} \|_{L_{\rho_X}^{p^*}} \le C_{p',\rho} \{ \varepsilon_\tau(f) - \varepsilon_\tau(f_{\rho,\tau}) \}^{\frac{1}{p'}}, \quad (8)$$

Where $p^* = \frac{pp'}{p+1} > 0$, and

$$C_{p',\rho} = 2^{1-\frac{1}{p'}} p'^{\frac{1}{p'}} \| \{ (b_x a_x^{p'-1})^{-1} \}_{x \in X} \|_{L_{\rho_X}^p}^{\frac{1}{p'}}.$$

By this conclusion, we can bound the error in learning theory scheme. Note that $\varepsilon_\tau(f) - \varepsilon_\tau(f_{\rho,\tau})$ depends on random samples $\{z_i = (x_i, y_i)\}_{i=1}^m$ and the ability of hypothesis space approximating the target function $f_{\rho,\tau}$, we define the regularization function f_λ by $f_\lambda = \arg\min_{f\in H_K}\{\varepsilon_\tau(f)-\varepsilon_\tau(f_{\rho,\tau})+\lambda\|f\|^2_{\hat{K}}\}$. (9)

Here $\hat{K}(x,y)=\int_X \tilde{K}(x,t)\tilde{K}(y,t)d\rho_X(t)$ and $\tilde{K}(x,y)=\int_X K(x,t)K(y,t)d\rho_X(t)$.

Although kernel K is not positive semi-definite, $\tilde{K}$ and $\hat{K}$ both are Mercer kernels, and associated reproducing kernel Hilbert spaces are denoted by $H_{\tilde{K}}$ and $H_{\hat{K}}$. Thus $L_{\tilde{K}}$ and $L_{\hat{K}}$ are compact, self-adjoint, positive operator on $L^2_{\rho_X}$, and there holds

$$L_{\tilde{K}} = L_K L_K^*, L_{\hat{K}} = L^2_{\tilde{K}}. \quad (10)$$

By the fact that $f_\lambda \in H_{\hat{K}}$, there exists $h_\lambda \in L^2_{\rho_X}$ such that $f_\lambda = L^{\frac{1}{2}}_{\hat{K}} h_\lambda = L_K L_K^* h_\lambda$. (11)

Through introducing continuous function $g_\lambda = L_K^* h_\lambda$, and

$$\hat{f}_{z,\lambda} = \frac{1}{m}\sum_{i=1}^m g_\lambda(x_i)K(\cdot,x_i) \in H_{K,z}.$$

we can decompose the excess generalization error as the following.

Proposition 1. Let λ>0 and $f^{(\varepsilon)}_{z,\lambda}=\sum_{i=1}^m \alpha_{z,i}K(\cdot,x_i)$ given by (5). Then

$$\varepsilon_\tau\left(\pi(f^{(\varepsilon)}_{z,\lambda})\right)-\varepsilon_\tau(f_{\rho,\tau})+\lambda\Omega(f^{(\varepsilon)}_{z,\lambda}) \le S_1+S_2+H_1+H_2+(1+\kappa^q)D(\lambda) \quad (12)$$

Where

$$S_1=\{\varepsilon_\tau(\pi(f^{(\varepsilon)}_{z,\lambda}))-\varepsilon_\tau(f_{\rho,\tau})\}-\{\varepsilon_{\tau,z}(\pi(f^{(\varepsilon)}_{z,\lambda}))-\varepsilon_{\tau,z}(f_{\rho,\tau})\},$$

$$S_2=\{\varepsilon_{\tau,z}(\pi(\hat{f}_{z,\lambda}))-\varepsilon_{\tau,z}(f_{\rho,\tau})\}-\{\varepsilon_\tau(\pi(\hat{f}_{z,\lambda}))-\varepsilon_\tau(f_{\rho,\tau})\}$$

$$H_1=\lambda\Omega(\hat{f}_{z,\lambda})-\lambda\|g_\lambda\|^q_{L^q_{\rho_X}}, H_2=\varepsilon_\tau(\hat{f}_{z,\lambda})-\varepsilon_\tau(f_\lambda),$$

$$D(\lambda)=\varepsilon_\tau(f_\lambda)-\varepsilon_\tau(f_{\rho,\tau})+\lambda\|f_\lambda\|^q_{\hat{K}}.$$

The left hand side of (12) can be decomposed as

$$\{\varepsilon_\tau\left(\pi(f^{(\varepsilon)}_{z,\lambda})\right)-\varepsilon_\tau(f_{\rho,\tau})\}-\{\varepsilon_{\tau,z}\left(\pi(f^{(\varepsilon)}_{z,\lambda})\right)-\varepsilon_{\tau,z}(f_{\rho,\tau})\}$$
$$+\{\varepsilon_{\tau,z}(\pi(f^{(\varepsilon)}_{z,\lambda}))+\lambda\Omega(\hat{f}^{(\varepsilon)}_{z,\lambda})\}-\{\varepsilon_{\tau,z}(\pi\hat{f}_{z,\lambda}))+\lambda\Omega(\hat{f}_{z,\lambda}\}$$
$$+\{\varepsilon_{\tau,z}(\hat{f}_{z,\lambda})-\varepsilon_{\tau,z}(f_{\rho,\tau})\}-\{\varepsilon_\tau(\hat{f}_{z,\lambda})-\varepsilon_\tau(f_{\rho,\tau})\}$$
$$+\{\varepsilon_\tau(\hat{f}_{z,\lambda})-\varepsilon_\tau(f_\lambda)\}+\{\varepsilon_\tau(f_\lambda)-\varepsilon_\tau(f_{\rho,\tau})\}+\lambda\|g_\lambda\|^q_{L^2_{\rho_X}}$$
$$+\{\lambda\Omega(\hat{f}_{z,\lambda})-\lambda\|g_\lambda\|^q_{L^q_{\rho_X}}\}+\{\lambda\|g_\lambda\|^q_{L^q_{\rho_X}}-\lambda\|g_\lambda\|^q_{L^2_{\rho_X}}\}$$

Observe that $\varepsilon_{\tau,z}\left(\pi(f^{(\varepsilon)}_{z,\lambda})\right)\le\varepsilon_{\tau,z}(f^{(\varepsilon)}_{z,\lambda})$ by |y|≤1, then the second item of the above equation is at most zero by the definition of $\hat{f}_{z,\lambda}$ and $\psi^{(\varepsilon)}_\tau(u)$. The last item is also at most zero due to $\|g_\lambda\|_{L^q_{\rho_X}}\le\|g_\lambda\|_{L^2_{\rho_X}}$ The fifth item is less than $(1+\kappa^q)D(\lambda)$ for the reason that $\|g_\lambda\|_{L^2_{\rho_X}}\le\|f_\lambda\|_{\tilde{K}}\le\kappa\|f_\lambda\|_{\hat{K}}$. This proves the desired inequality.

Let $\lambda_1\ge\lambda_2\ge\cdots$ be the nonnegative eigenvalues of integral operator $L_{\tilde{K}}$, and $\varphi_k, k\in N$ be the associated orthonormal eigenfunctions.

Our approximation condition is given as

$$f_{\rho,\tau}=L^r_{\tilde{K}}g_{\rho,\tau}, \text{for some } 0<r\le 1, g_{\rho,\tau}\in L^2_{\rho_X}(X). \quad (13)$$

Proposition 2. Under the approximation condition (13), and 0<λ≤1 . Then there holds $D(\lambda)\le C_0\lambda^{\frac{r}{r+(1-r)q}}$ with the constant $C_0=\|g_{\rho,\tau}\|_{L^2_{\rho_X}}+\|g_{\rho,\tau}\|^q_{L^2_{\rho_X}}$.

Proof. Because ψ_τ is a Lipschitz function, $D(\lambda)\le\|f_\lambda-f_{\rho,\tau}\|_{L^2_{\rho_X}}+\lambda\|f_\lambda\|^q_{\hat{K}}$.

It follows from (13), $f_{\rho,\tau}=\sum_{k=1}^\infty\lambda_k^r\alpha_k\varphi_k$ when $g_{\rho,\tau}=\sum_{k=1}^\infty\alpha_k\varphi_k$ .Let h=r+(1-r)q. When $0<\lambda\le\lambda_1^h$, there exists N, such that $\lambda_{N+1}<\lambda^{\frac{1}{h}}\le\lambda_N$. Taking $f_N=\sum_{k=1}^N\lambda_k^r\alpha_k\varphi_k$, we have

$$\|f_N\|^2_{\hat{K}}=\sum_{k=1}^N\lambda_k^{2r-2}\alpha_k^2\le\lambda_k^{2r-2}\|g_{\rho,\tau}\|^2_{L^2_{\rho_X}}\le\lambda^{\frac{2r-2}{h}}\|g_{\rho,\tau}\|^2_{L^2_{\rho_X}}$$

and

$$\|f_N-f_{\rho,\tau}\|^2_{L^2_{\rho_X}}=\|\sum_{k=N+1}^\infty\lambda_k^r\alpha_k\varphi_k\|^2_{L^2_{\rho_X}}\le\lambda^{\frac{2r}{h}}\|g_{\rho,\tau}\|^2_{L^2_{\rho_X}}$$

Then

$$D(\lambda)\le\|f_N-f_{\rho,\tau}\|_{L^2_{\rho_X}}+\lambda\|f_N\|^q_{\hat{K}}$$
$$\le(\|g_{\rho,\tau}\|_{L^2_{\rho_X}}+\|g_{\rho,\tau}\|^q_{L^2_{\rho_X}})\lambda^{\frac{r}{r+(1-r)q}}$$

For the case $\lambda>\lambda_1^h$, the conclusion holds by $D(\lambda)\le\|f_{\rho,\tau}\|_{L^2_{\rho_X}}$. The proof is complete.

Our sample error is estimated through a concentration inequality, so the capacity of the hypothesis space plays an important role. Covering numbers are often utilized to measure the capacity which have been well studied in [21] .

Let *F* be a set of functions on X, $x=\{x_i\}_{i=1}^k\in X^k$. The sampling operator $S_x: F\to\Re^k$, is defined by

$S_x(f)=\{f(x_i)\}_{i=1}^k$, for $f \in \mathcal{F}$. The l^2-empirical covering number of $\mathcal{F}$ is defined by

$$N_2(\mathcal{F},\varepsilon)=\sup_{k\in N}\sup_{x\in X^k} N_2(S_x(\mathcal{F}),\varepsilon) \quad (14)$$

Here $N_2(S_x(\mathcal{F}),\varepsilon)$ is the covering number of $S_x(\mathcal{F})$ in Euclidean space $\Re^k$ with the normalized l^2-metric d_2.

For any R>0, denote that

$$B_R=\{f\in H_{K,z}, \|f\|_z:=(m^{q-1}\sum_{i=1}^m|\alpha_i|^q)^{\frac{1}{q}}\le R\}. (15)$$

It is proved in [7] that if $K\in C^s(X\times X)$, there holds

$$\log N_2(B_1,\varepsilon)\le c_{\mu,K}\varepsilon^{-\mu}, \forall\varepsilon>0. \quad (16)$$

Here $c_{\mu,K}$ is a constant independent of ε>0 and $0<\mu<2$ is a power index defined by

$$\mu=\begin{cases}2n/(n+2s), & \text{when } 0<s\le 1,\\ 2n/(n+2), & \text{when } 1<s\le 1+n/2,\\ n/s, & \text{when } s>1+n/2.\end{cases}$$

Our main result is given in the following that will be proved in Section 4.

Theorem 1. Assume approximation condition (13) with 0<r≤1 and capacity condition (16) with 0<μ<2 hold. Suppose that ρ has a τ-quantile of p-average type p' for some $p\in(0,+\infty]$ and $p'\in(1,+\infty)$, $p^*=\frac{pp'}{p+1}>0$. Taking $\lambda=m^{-\frac{(1-q)r+q}{2}}$, and $\varepsilon=m^{\omega}$ with $\frac{r}{2}\le\omega<\omega_0$. Then, for any 0<δ<1, with confidence 1-δ, we have

$$\|\pi(f_{z,\lambda}^{(\varepsilon)})-f_{\rho,\tau}\|_{L_{\rho_X}^{p^*}}^{p'}\le b_0\{b(\theta,\mu,\frac{\delta}{2},\eta)^{\frac{2\mu}{2+\mu}} + (1+\frac{1}{m}\log\frac{20}{\delta})^{\frac{1}{q}}\log\frac{20}{\delta}\}m^{-\frac{r}{2}} \quad (17)$$

Here b_0 is a constant independent of m or δ, and $b(\theta,\mu,\frac{\delta}{2},\eta)$ is given by (25).

We deduce the same learning rate $O(m^{-\frac{r}{2}})$ as that for the quantile regression learning scheme with ε-insensitive pinball loss and l^1-norm regularization in Ref.[5]. But the constant $b(\theta,\mu,\frac{\delta}{2},\eta)$ and the term $(1+\frac{1}{m}\log\frac{20}{\delta})^{\frac{1}{q}}$ in the right hand side of (17) is related to regularization parameter q.

3 Estimates for hypothesis errors and sample errors

In this section, we firstly estimate the hypothesis error H_1 and H_2 by the following probability inequality in [9].

Lemma 1. Let H be a Hilbert space and ξ be a random variable on a probability space (Z,ρ) with values in H. Assume $\|\xi\|\le\tilde{M}<\infty$ almost surely. Denote $\sigma^2(\xi)=E(\|\xi\|^2)$. Let $\{\xi_i\}_{i=1}^m$ be independent random drawers of ξ. For any 0<δ<1, with confidence 1-δ,

$$\|\frac{1}{m}\sum_{i=1}^m[\xi_i-E(\xi_i)]\|\le\frac{2\tilde{M}\log\frac{2}{\delta}}{m}+\sqrt{\frac{2\sigma^2(\xi)\log\frac{2}{\delta}}{m}}.$$

Proposition 3. For any 0<δ<1, with confidence 1-δ, we have

$$H_1\le\frac{3\kappa^q D(\lambda)\log\frac{4}{\delta}}{m}+\frac{\kappa^q}{2}D(\lambda),$$

and

$$H_2\le\kappa^2\frac{D(\lambda)^{\frac{1}{q}}}{\lambda^{\frac{1}{q}}\sqrt{m}}\{\frac{2\log\frac{4}{\delta}}{\sqrt{m}}+\sqrt{2\log\frac{4}{\delta}}\}.$$

Proof. Let us deal with H_1 first. Consider the random variable $\xi=|g_\lambda(x)|^q$ on (X,ρ_X) with values in $\Re$, then

$$|g_\lambda(x)|^q=|\int_X K(t,x)h_\lambda(t)d\rho_X(t)|^q\le\kappa^q\|h_\lambda\|_{L_{\rho_X}^q}^q$$
$$\le\kappa^q\|h_\lambda\|_{L_{\rho_X}^2}^q=\kappa^q\|f_\lambda\|_K^q\le\kappa^q\frac{D(\lambda)}{\lambda}$$

Hence, $|\xi|\le\kappa^q\frac{D(\lambda)}{\lambda}$ .It easy to see that $E\xi=\|g_\lambda\|_{L_{\rho_X}^q}^q$, and

$$\sigma^2(\xi)\le E(\xi^2)=\int_X|g_\lambda|^{2q}d\rho_X\le\kappa^{2q}\frac{D^2(\lambda)}{\lambda^2}.$$

Applying Lemma 1 to $\xi=|g_\lambda(x)|^q$, we have with confidence 1-δ/2,

$$H_1=\lambda\Omega(\hat{f}_{z,\lambda})-\lambda\|g_\lambda\|_{L_{\rho_X}^q}^q$$
$$\le\frac{2\kappa^q D(\lambda)}{m}\log\frac{4}{\delta}+\sqrt{\frac{2\kappa^{2q}D^2(\lambda)}{m}\log\frac{4}{\delta}}$$
$$\le\frac{3\kappa^q D(\lambda)}{m}\log\frac{4}{\delta}+\frac{\kappa^q}{2}D(\lambda).$$

As for H_2. Since ψ_τ satisfies Lipschitz condition,

$$\varepsilon_\tau(\hat{f}_{z,\lambda})-\varepsilon_\tau(f_\lambda) \le \int_X |\hat{f}_{z,\lambda}(x)-f_\lambda(x)|\,d\rho_X(x) \le \|\hat{f}_{z,\lambda}-f_\lambda\|_{L^2_{\rho_X}} \quad (18)$$

Apply Lemma 1 to the random variables $\zeta(x)=g_\lambda(x)K_x$ on (X,ρ_X) with values in the Hilbert space $L^2_{\rho_X}$. It satisfies $E(\zeta)=L_K g_\lambda = f_\lambda$,

$$\|\zeta\|_{L^2_{\rho_X}} \le \kappa\|g_\lambda\|_\infty \le \kappa^2\left(\frac{D(\lambda)}{\lambda}\right)^{\frac{1}{q}},$$

$\sigma^2(\zeta)\le\kappa^4\left(\frac{D(\lambda)}{\lambda}\right)^{\frac{2}{q}}$. Thus with confidence $1-\frac{\delta}{2}$, we obtain,

$$\|\hat{f}_{z,\lambda}-f_\lambda\|_{L^2_{\rho_X}} \le \frac{2\kappa^2 D(\lambda)^{\frac{1}{q}}}{m\lambda^{\frac{1}{q}}}\log\frac{4}{\delta}+\sqrt{\frac{2\kappa^4 D(\lambda)^{\frac{2}{q}}}{m\lambda^{\frac{2}{q}}}\log\frac{4}{\delta}}$$

$$=\kappa^2\frac{D(\lambda)^{\frac{1}{q}}}{m^{\frac{1}{2}}\lambda^{\frac{1}{q}}}\{2m^{-\frac{1}{2}}\log\frac{4}{\delta}+\sqrt{2\log\frac{4}{\delta}}\}.$$

This implies the desired estimate.
Let θ be given by

$$\theta=\min\{\frac{2}{p'},\frac{p}{p+1}\}\in(0,1]. \quad (19)$$

and the constant $C_\theta = 2^{2-\theta}p'^{\theta}\|\gamma^{-1}\|^{\theta}_{L^p_{\rho_X}}$. The estimate of sample error S_2 follows.

Proposition 4. Suppose that ρ has a τ-quantile of p-average type p' for some $p\in(0,+\infty]$ and $p'\in(1,+\infty)$. Assume B_1 satisfies (16) with some $0<\mu<2$. Let $R\ge1$ and $0<\lambda\le1$. Then, for any $0<\delta<1$, with confidence $1-\delta$, we have

$$S_2\le C_2\left(1+\frac{1}{m}\log\frac{5}{\delta}\right)^{\frac{1}{q}}\log\frac{5}{\delta}\times m^{-\frac{2(1-\theta)}{4-2\theta+\mu\theta}}$$

$$\times\left(\lambda^{\frac{r-1}{(1-q)r+q}}m^{-\frac{\theta}{2}}+\lambda^{\frac{r-1+\theta}{(1-q)r+q}}\right).$$

Here C_2 is a constant independent of m,λ,δ .

Proof. Recall the estimate of H_1, for any $0<\delta<1$, with confidence $1-2\delta/5$, there holds

$$\frac{1}{m}\sum_{i=1}^m|g_\lambda(x_i)|^q-\|g_\lambda\|^q_{L^q_{\rho_X}}\le\frac{3\kappa^q D(\lambda)}{m\lambda}\log\frac{5}{\delta}+\frac{\kappa^q D(\lambda)}{2\lambda},$$

which implies the existence of a subset $U_1\subset Z^m$ with measure at most $2\delta/5$, such that

$$\frac{1}{m}\sum_{i=1}^m|g_\lambda(x_i)|^q\le\max\{\frac{3\kappa^q D(\lambda)}{m\lambda}\log\frac{5}{\delta}+\frac{3\kappa^q D(\lambda)}{2\lambda},1\}$$

$$\overset{\Delta}{=}R^q_\lambda,\quad \forall z\in Z^m\setminus U_1.$$

This inequality ensures that for every $z\in Z^m\setminus U_1$, we have $\hat{f}_{z,\lambda}\in B_{R_\lambda}$.

By the proof of Proposition 5 in [4], there exists U_{R_λ} with measure at most $\delta/5$ such that for every $z\in Z^m\setminus(U_1\bigcup U_{R_\lambda})$, we have

$$\{\varepsilon_{\tau,z}(\hat{f}_{z,\lambda})-\varepsilon_{\tau,z}(f_{\rho,\tau})\}-\{\varepsilon_\tau(\hat{f}_{z,\lambda})-\varepsilon_\tau(f_{\rho,\tau})\}$$

$$\le\frac{1}{2}\tilde{C}^{1-\theta}R_\lambda^{1-\theta}m^{-\frac{2(1-\theta)}{4-(2-\mu)\theta}}|\varepsilon_\tau(\hat{f}_{z,\lambda})-\varepsilon_\tau(f_\lambda)|^\theta$$

$$+\frac{1}{2}\tilde{C}^{1-\theta}R_\lambda^{1-\theta}m^{-\frac{2(1-\theta)}{4-(2-\mu)\theta}}|\varepsilon_\tau(f_\lambda)-\varepsilon_\tau(f_{\rho,\tau})|^\theta \quad (20)$$

$$+2\left(\frac{c_\theta}{m}\log\frac{5}{\delta}\right)^{\frac{1}{2-\theta}}+c_\mu\tilde{C}R_\lambda m^{-\frac{2}{4-(2-\mu)\theta}}$$

$$+18(\kappa+1)\frac{R_\lambda}{m}\log\frac{5}{\delta}\max\{R,1\}.$$

Proposition 2 yields that

$$R_\lambda\le((3C_0)^{\frac{1}{q}}\kappa+1)\lambda^{\frac{r-1}{(1-q)r+q}}\left(\frac{1}{m}\log\frac{5}{\delta}+1\right)^{\frac{1}{q}},$$

and

$$\varepsilon_\tau(f_\lambda)-\varepsilon_\tau(f_{\rho,\tau})\le D(\lambda)\le C_0\lambda^{\frac{r-1}{(1-q)r+q}}.$$

Moreover, from Proposition 3, we know that there exists a subset U_2 with measure at most $2\delta/5$ such that for every $z\in Z^m\setminus U_2$,

$$\varepsilon_\tau(\hat{f}_{z,\lambda})-\varepsilon_\tau(f_\lambda)\le\kappa^2\frac{D(\lambda)^{\frac{1}{q}}}{m^{\frac{1}{2}}\lambda^{\frac{1}{q}}}\{m^{-\frac{1}{2}}\log\frac{5}{\delta}+\sqrt{2\log\frac{5}{\delta}}\}.$$

Finally, let $U=U_1\bigcup U_2\bigcup U_{R_\lambda}$, the measure of U is at most δ. For every $z\in Z^m\setminus U$, by plugging the above estimates into (20), Proposition 4 holds.

We next bound S_1 by Proposition 5 proved in [4].

Proposition 5. Under assumptions of Proposition 4. Let $R\ge1$ and $0<\lambda\le1$. Then, for any $0V\delta<1$ and all $f\in B_R$, with confidence $1-\delta$, there holds

$$\{\varepsilon_\tau(\pi(f))-\varepsilon_\tau(f_{\rho,\tau})\}-\{\varepsilon_{\tau,z}(\pi(f))-\varepsilon_{\tau,z}(f_{\rho,\tau})\}$$

$$\le\frac{1}{2}C_1^{1-\theta}R^{\frac{2\mu\mu(-\theta)}{2+\mu}}m^{-\frac{2(1-\theta)}{4+\mu\theta-2\theta}}\{\varepsilon_\tau(\pi(f))-\varepsilon_\tau(f_{\rho,\tau})\}^\theta$$

$$+(36+2C_\theta^{\frac{1}{2-\theta}})\log\frac{1}{\delta}m^{-\frac{1}{2-\theta}}+c_\mu C_1 R^{\frac{2\mu}{2+\mu}}m^{-\frac{2}{4+\mu\theta-2\theta}}$$

Here C_1 is the constant dependent on the constants $\mu,\theta,c_{\mu,K},C_\theta$.

4 Error Bound and Convergence Rates by Iteration

In this section, we deduce the error bound and convergence rate by the iteration technique.

For R≥1, denote $w(R)=\{z\in Z^m:\|f_{z,\lambda}^{(\varepsilon)}\|\le R\}$.

Proposition 6. Assume that (13) and (16) hold. Suppose that ρ has a τ-quantile of p-average type p' for some $p\in(0,\infty]$ and $p'\in(1,\infty)$. Let 0<λ≤1,R≥1, and 0<δ<1. Then, there exists a subset V_R of Z^m with measure at most δ such that for any $z\in W(R)\setminus V_R$,

$$\varepsilon_\tau(\pi(f_{z,\lambda}^{(\varepsilon)}))-\varepsilon_\tau(f_{\rho,\tau})+\lambda\Omega(f_{z,\lambda}^{(\varepsilon)})$$
$$\le\hat{C}R^{\frac{2\mu}{2+\mu}}m^{-\frac{2}{4+\mu\theta-2\theta}}+C_3(1+\frac{1}{m}\log\frac{10}{\delta})\log\frac{10}{\delta}\Phi(m,\lambda)+2\varepsilon$$

Here $\hat{C}$ and C_3 are constants independent of m,λ,δ, and

$$\Phi(m,\lambda)=\lambda^{\frac{r}{(1-q)r+q}}+\lambda^{\frac{r-1}{(1-q)r+q}}m^{-\frac{1}{2}}+\lambda^{\frac{r-1-\theta}{(1-q)r+q}}m^{-\frac{2(1-\theta)}{4-2\theta+\mu\theta}}.\ (21)$$

From proposition 3 we know that there exists V_1 of Z^m with measure at most 2δ/5 such that

$$H_1\le\frac{3\kappa^qD(\lambda)}{m}\log\frac{10}{\delta}+\frac{\kappa^q}{2}D(\lambda)$$

and

$$H_2\le\kappa^2\frac{D(\lambda)^{\frac{1}{q}}}{\sqrt{m}\lambda^{\frac{1}{q}}}\{\sqrt{m}\log\frac{4}{\delta}+\sqrt{2\log\frac{4}{\delta}}\}$$

hold for any $z\in Z^m\setminus V_1$.

From proposition 4 we know that there exists a subset U_R of Z^m with measure at most δ/10, such that for $\forall z\in W(R)\setminus U_R$

$$S_1\le\frac{1}{2}C_1^{1-\theta}R^{\frac{2\mu(1-\theta)}{2+\mu}}m^{-\frac{2(1-\theta)}{4+\mu\theta-2\theta}}\{\varepsilon_\tau(\pi(f_{z,\lambda}^{(\varepsilon)}))-\varepsilon_\tau(f_{\rho,\tau})\}^\theta$$
$$+(36+2C_\theta^{\frac{1}{2-\theta}})\log\frac{10}{\delta}m^{-\frac{1}{2-\theta}}+c_\mu C_1R^{\frac{2\mu}{2+\mu}}m^{-\frac{2}{4+\mu\theta-2\theta}}.$$

From proposition 5 we know that there exists a subset V_2 of Z^m with measure at most δ/2, such that for $\forall z\in Z^m\setminus V_2$,

$$S_2\le C_2(1+\frac{1}{m}\log\frac{10}{\delta})^{\frac{1}{q}}\log\frac{10}{\delta}\times m^{-\frac{2(1-\theta)}{4-2\theta+\mu\theta}}$$
$$\times(\lambda^{\frac{r-1}{(1-q)r+q}}m^{-\frac{\theta}{2}}+\lambda^{\frac{r-1+\theta}{(1-q)r+q}}).$$

Thus,by taking $V_R=V_1\bigcup V_2\bigcup V_R$, the measure of V_R is at most δ. With the above bounds combined in proposition 1, for every $\forall z\in W(R)\setminus V_R$, there holds

$$\varepsilon_\tau(\pi(f_{z,\lambda}^{(\varepsilon)}))-\varepsilon_\tau(f_{\rho,\tau})+\lambda\Omega(f_{z,\lambda}^{(\varepsilon)})\le\frac{1}{2}C_3(1+\frac{1}{m}\log\frac{10}{\delta})$$
$$\times\log\frac{10}{\delta}\Phi(m,\lambda)+\frac{1}{2}C_1^{1-\theta}R^{\frac{2\mu(1-\theta)}{2+\mu}}m^{-\frac{2(1-\theta)}{4+\mu\theta-2\theta}}$$
$$\times\{\varepsilon_\tau(\pi(f_{z,\lambda}^{(\varepsilon)}))-\varepsilon_\tau(f_{\rho,\tau})\}^\theta+c_\mu C_1R^{\frac{2\mu}{2+\mu}}m^{-\frac{2}{4+\mu\theta-2\theta}}+\varepsilon.$$

Let $t=\varepsilon_\tau(\pi(f_{z,\lambda}^{(\varepsilon)}))-\varepsilon_\tau(f_{\rho,\tau})+\lambda\Omega(f_{z,\lambda}^{(\varepsilon)})$. The above formula can be expressed as

$$t-\frac{1}{2}C_1^{1-\theta}R^{\frac{2\mu(1-\theta)}{2+\mu}}m^{-\frac{2(1-\theta)}{4+\mu\theta-2\theta}}t^\theta-\Pi\le0.\qquad(22)$$

Where Π is the rest three terms. By Lemma 7.2 in [3], the unique positive solution t^* can be bounded as

$$t^*\le\max\{C_1R^{\frac{2\mu(1-\theta)}{2+\mu}}m^{-\frac{2(1-\theta)}{4+\mu\theta-2\theta}},2\Pi\}$$
$$\le C_1R^{\frac{2\mu(1-\theta)}{2+\mu}}m^{-\frac{2(1-\theta)}{4+\mu\theta-2\theta}}+2\Pi$$

Hence, we complete the proof.

Lemma 2. For almost every $z\in Z^m$, we have

$$\|f_{z,\lambda}^{(\varepsilon)}\|_z\le(\frac{1}{\lambda})^{\frac{1}{q}}.\qquad(23)$$

By the definition of $f_{z,\lambda}$ and |y|≤1 ,we have

$$\lambda\|f_{z,\lambda}^{(\varepsilon)}\|_z^q\le\frac{1}{m}\sum_{i=1}^m\psi_\tau^{(\varepsilon)}(f_{z,\lambda}^{(\varepsilon)}(x_i)-y_i)+\lambda\Omega(f_{z,\lambda}^{(\varepsilon)})$$
$$\le\frac{1}{m}\sum_{i=1}^m\psi_\tau^{(\varepsilon)}(-y_i)+\lambda\Omega(0)\le1$$

So **Lemma 2** holds almost surely.

The above bound for $\|f_{z,\lambda}^{(\varepsilon)}\|_z$ is rough, to obtain a tighter bound we shall apply iteration technique that has been widely used in learning error estimate, see [11,16]. For simplicity we denote $\omega_0=\dfrac{2}{4-w\theta+\mu\theta}$ in the following.

Lemma 3. Under the assumptions in proposition 6, taking $\lambda=m^{-\beta}$ and $\varepsilon=m^{-\omega}$ with $0<\beta,\omega\le1/2$. Suppose that ρ has a τ-quantile of p-average type p' for some $p\in(0,+\infty]$ and $p'\in(1,+\infty)$. Then, for any 0<δ<1 and arbitrarily small 0<η≤1, with confidence 1-δ, there holds

$$\|f_{z,\lambda}^{(\varepsilon)}\|_z\le(M^{\frac{\mu}{q^2(2+\mu)}}(1+C_q)^{\frac{2+\mu}{q(2+\mu)-\mu}}+\overline{C})b(\theta,\mu,\delta,\eta)m^{\theta\eta}.(24)$$

Where

$$b(\theta,\mu,\delta,\eta)=(1+L_{\theta,\mu,\eta})(\log\frac{10}{\delta}+\log L_{\theta,\mu,\eta})^{\frac{1}{q}}$$

$$\times(1+\frac{1}{m}(\log\frac{10}{\delta}+\log L_{\theta,\mu,\eta}))^{\frac{1}{q}} \qquad (25)$$

and $L_{\theta,\mu,\eta}$ is given by (31), θ_η is given by

$$\theta_\eta=\max\frac{1}{q}\{\beta(1-\frac{r}{(1-q)r+q})+\max\{0,\frac{\beta}{(1-q)r+q}-\frac{1}{2},$$
$$\frac{(1-\theta)\beta}{(1-q)r+q}-(1-\theta)\omega_0\},\beta-\omega,(\beta-\omega_0)\frac{1}{1-\Delta}+\eta\}.$$

Proof. Taking $\lambda=m^{-\beta}$ and $\varepsilon=m^{-\omega}$ in proposition 6, for any R≥1 there exists a subset $V_R\subset Z^m$ with measure at most δ such that

$$\|f_{z,\lambda}^{(\varepsilon)}\|_z\le a_m R^{\frac{2\mu}{q(2+\mu)}}+b_m,\quad \forall z\in W(R)\setminus V_R. \quad (26)$$

Where

$$a_m=C_q m^{\frac{\beta-\omega_0}{q}},$$

$$b_m=\overline{C}(1+\frac{1}{m}\log\frac{10}{\delta})^{\frac{1}{q}}(\log\frac{10}{\delta})^{\frac{1}{q}}m^{\gamma}\overset{\Delta}{=}b_\delta m^{\gamma}.$$

Here $C_q,\overline{C}$ are constants independent of m,λ,δ and

$$\gamma=\max\frac{1}{q}\{\beta(1-\frac{r}{(1-q)r+q})+\max\{0,\frac{\beta}{(1-q)r+q}-\frac{1}{2},$$
$$\frac{(1-\theta)\beta}{(1-q)r+q}-(1-\theta)\omega_0\},\beta-\omega\}.$$

It follows that

$$W(R)\subset W(a_m R^{\frac{2\mu}{q(2+\mu)}}+b_m)\cup V_R. \qquad (27)$$

Let us define a sequence $\{R^{(l)}\}_{l=0}^{L}$, $R^{(0)}=(\frac{1}{\lambda})^{\frac{1}{q}}$ and

$$R^{(l)}=a_m(R^{(l-1)})^{\frac{2\mu}{q(2+\mu)}}+b_m,\quad l\in N \qquad (28)$$

Lemma 2 gives the identity $W(R^{(0)})=Z^m$. We see that

$$Z^m=W(R^{(0)})\subseteq W(R^{(1)})\cup V_{R^{(0)}}\subseteq\cdots W(R^{(L)})\cup(\bigcup_{l=0}^{L-1}V_{R^{(l)}}).$$

Since the measure of $V_{R^{(l)}}$ is at most δ, the measure of $W(R^{(L)})$ is at least 1-Lδ. Denote $\Delta=\frac{2\mu}{q(2+\mu)}<1$, we have

$$R^{(L)}\le a_m^{1+\Delta+\Delta^2+\cdots+\Delta^{L-1}}(R^{(0)})^{\Delta^L}+\sum_{l=1}^{L-1}a_m^{1+\Delta+\Delta^2+\cdots+\Delta^{l-1}}b_m^{\Delta^{l-1}}+b_m$$

$$\le M^{\frac{\Delta}{q}}(1+C_q)^{\frac{1}{q(1-\Delta)}}m^{\frac{1}{q}\{(\beta-\omega_0)\frac{1}{1-\Delta}+\Delta^L(\beta-(\frac{1}{1-\Delta}(\beta-\omega_0)))\}}$$
$$+a_m^{\frac{1}{1-\Delta}}\sum_{l=1}^{L-1}(a_m^{-\frac{1}{1-\Delta}}b_m)^{\Delta^l}+b_m$$

$$\le M^{\frac{\mu}{q^2(2+\mu)}}(1+C_q)^{\frac{2+\mu}{q(2+\mu)-2\mu}}$$
$$\times m^{\frac{1}{q}\{(\beta-\omega_0)\frac{q(2+\mu)}{q(2+\mu)-2\mu}+\Delta^L(\frac{q(2+\mu)}{q(2+\mu)-2\mu}\omega_0+\frac{q(2+\mu)-3\mu-2}{q(2+\mu)-\mu}\beta)\}}$$
$$+(L+1)b_\delta m^{\gamma}+LC_q^{\frac{q(2+\mu)}{q(2+\mu)-2\mu}}m^{(\beta-\omega_0)\frac{q(2+\mu)}{q(2+\mu)-2\mu}}$$

When $\beta<\frac{\omega_0}{\Delta}$, let

$$L_\eta=\max\{(\log_{\frac{q(2+\mu)}{2\mu}}\frac{\omega_0-\Delta\beta}{(1-\Delta)\eta})+1,1\}. \qquad (29)$$

Otherwise, let $L_\eta=1$. We have that

$$\Delta^L(\beta-\frac{1}{1-\Delta}(\beta-\omega_0))\le\eta.$$

Then, with confidence $1-L_\eta\delta$, there holds

$$\|f_{z,\lambda}^{(\varepsilon)}\|_z\le(b_\delta+M^{\frac{\mu}{q^2(2+\mu)}}(1+C_q)^{\frac{2+\mu}{q(2+\mu)-\mu}})(1+L_\eta)m^{\theta_\eta}. \quad (30)$$

By a simple calculation we can obtain

$$L_\eta\le\log_{\frac{q(2+\mu)}{2\mu}}\frac{\omega_0}{(1-\Delta)\eta}+1\overset{\Delta}{=}L_{\theta,\mu,\eta}. \qquad (31)$$

Then our result follows by replacing δ as $\delta/L_{\theta,\mu,\eta}$.

Finally we are in the position to prove our main results.

Combine Lemma 3 with Proposition6, and replace δ by δ/2 in both results, with confidence 1-δ, we have

$$\|\pi(f_{z,\lambda}^{(\varepsilon)})-f_{\rho,\tau}\|_{L_{\rho X}^{p_*}}^{p'}\le b_1\{b(\theta,\mu,\frac{\delta}{2},\eta)^{\frac{2\mu}{2+\mu}}$$

$$+(1+\frac{1}{m}\log\frac{20}{\delta})^{\frac{1}{q}}\log\frac{20}{\delta})\Psi(m,\lambda) \qquad (32)$$

Here b_1 is a constant independent of m , δ and η, and

$$\Psi(m,\lambda)=\Phi(m,\lambda)+m^{\frac{2\mu}{2+\mu}\theta_\eta-\omega_0}+m^{-\omega}=O(m^{-\vartheta(\beta,\omega)}). \quad (33)$$

By choosing $\beta=\beta_0=\frac{(1-q)r+q}{2}$,

$0<\eta\le\frac{(1-r)q}{2}+\frac{\omega_0}{1-\Delta}-\frac{(1-q)r+q}{2(1-\Delta)}$, we have

$$\theta_\eta=\max\{\frac{1-r}{2},\frac{(1-q)r+q-2\omega}{2q}\}.$$

Thus

$$\vartheta(\beta,\omega)=\min\{\omega,\frac{r}{2},\omega_0-\frac{\mu(1-r)}{2+\mu},$$

$$\omega_0-\frac{\mu}{2+\mu}\frac{(1-q)r+q-2\omega}{q}\}$$

Our conclusion follows by taking $\frac{r}{2}\leq\omega<\omega_0$. The proof of Theorem1 is complete.

Acknowledgments

The work described in this paper was supported by National Nature Science Foundation of China (Grants No 11671171).

References

1. N. Aronszajn, Theory of reproducing kernels, Trans. Amer. Math. Soc., 68 (1950), 337-404.
2. D. R. Chen, Q. Wu, Y. Ying, and D. X. Zhou, Support vector machine soft margin classifiers: error analysis, J. Machine Learning Research 5 (2004), 1143–1175.
3. F. Cucker and D. X. Zhou, Learning Theory: an Approximation Theory Viewpoint, Cambridge University Press, 2007.
4. M. Li, H. W. Sun, Asymptotic analysis of quantile regression learning based on coefficient dependent regularization, International Journal of Wavelets, Multiresolution and Information Processing,vol.13,no.4,2015.
5. M. Li, M. J. Zhang, H. W. Sun,Conditional quantile regression with l^1-regularization and ε-insensitive pinball loss, Biomedical Engineering and Informatics (BMEI),vol.8 pp.843-851,2015.
6. S. G. Lv, D. M. Shi, Q. W. Xiao and M. S. Zhang, Sharp learning rates of coefficient-based l^q-regularized regression with indefinite kernels, Sci China Math 56 (2013), 1557–1574.
7. L. Shi, Learning theory estimates for coefficient-based regularized regression, Applied and Computational Harmonic Analysis 34 (2013), 252–265.
8. L. Shi, Y. L. Feng and D. X. Zhou, Concentration estimates for learning with l^1-regularizer and data dependent hypothesis spaces, Applied and Computational Harmonic Analysis 31 (2011), 286–302.
9. S. Smale and D. X. Zhou, Learning theory estimates via integral operators and their approximations, Constructive Approximation. 26 (2007), 153–172.
10. A. Christmann and I. Steinwart, How SVMs can estimate quantile and the median, Advances in Neural Information Processing Systems 20 (2008), 305–312.
11. I. Steinwart and C. Scovel, Fast rates for support vector machines using Gaussian kernels, the Annals of Statistics. 35 (2007), 575–607.
12. I. Steinwart and A. Christmann, Estimating conditional quantiles with the help of the pinball loss, Bernoulli 17 (2011), 211–225.
13. H. Sun, Q. Wu, Sparse Representation in Kernel Machines. IEEE Transaction on Neural Networks and Learning Systems, 26(10): 2576-2582, 2015.
14. H. W. Sun and Q. Wu, Least square regression with indefinite kernel and coefficient regularization, Applied and Computational Harmonic Analysis 30 (2011), 96–109.
15. V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
16. Q. Wu, Y. Ying, D. X. Zhou, Learning rates of least-square regularized regression, Foundations of Computational Mathematics 6 (2006), 171–192.
17. Q. Wu and D. X. Zhou, Learning with sample dependent hypothesis spaces, Computers and Mathematics with Applications 56 (2008), 2896–2907.
18. D. H. Xiang, Conditional quantiles with varying Gaussian, Advances in Computational Mathematics 38(4) (2013) 723–735.
19. D. H. Xiang, T. Hu and D. X. Zhou, Approximation analysis of learning algorithms for support vector regression and quantile regression, Journal of Applied Mathematics, 2012 (2012), 1–17.
20. Q. W. Xiao, and D. X. Zhou. Learning by nonsymmetric kernels with data dependent spaces and 1-regularizer. Taiwanese Journal of Mathematics, 14(5): 1821-1836, 2010.
21. D. X. Zhou, The covering number in learning theory, Journal of Complexity 18 (2002), 739–767.